Intelligent Systems Group
University of Bristol

Additional material for the Ukwabelana Zulu corpus

Sebastian Spiegler[1], **Andrew van der Spuy**[2], **Peter A. Flach**[1]

[1] Intelligent Systems Group
University of Bristol
Bristol, United Kingdom

[2] School of Literature and Language Studies
University of the Witwatersrand
Johannesburg, South Africa

Technical Report CSTR-10-003

July 2010

Abstract

In this document we describe the scheme used for labelling the open-source Ukwabelana Zulu corpus as well as the rules employed for the Part-of-speech (POS) tagger used to assign POS to morphologically analysed words. A detailed description of the Zulu morphology, the corpus itself and its generation is given in [2]. All resources can be downloaded from `http://www.cs.bris.ac.uk/Research/MachineLearning/Morphology/Resources/`.

1 The labelling scheme of the Ukwabelana Zulu corpus

In this section we give describe the labelling scheme of the Ukwabelena corpus by listing all labels used with a short description, its frequency in the corpus and an example.

Label	Description	Freq.	Example
<adv>	adverb	351	a<i6>be<past>njalo<adv>
<advpf>	adverb prefix	38	i<der>ka<advpf>khulu<ar>
<ar>	adjective root	382	a<i6>luhlaza<ar>
<asp>	aspect	2914	a<i6>ka<asp>gang<vr>ile<vpl>
<asp_vr>	aspect + verb root	2	w<i1>ez<asp_vr>a<va>
<aug>	augmentative	3	y<p>i<iv>n<n9>dilinga<nr>kazi<aug>
<cj>	conjunction	21	k<z15>o<iv>kuba<cj>
<d>	demonstrative	147	a<z6>nga<p>le<d>s<d7>o<pos2>si<n7>khathi<nr>
<d1>	demonstrative agreement class 1	27	ba<i2>ku<locpf>lo<d1>
<d2>	demonstrative agreement class 2	17	ba<z2>la<d>ba<d2>
<d3>	demonstrative agreement class 3	17	b<z14>a<z6>ku<locpf>lo<d3>wa<st>ya<pos3>
<d4>	demonstrative agreement class 4	9	le<d4>na<st>
<d5>	demonstrative agreement class 5	17	e<p1>ku<locpf>le<d>li<d5>buth<vr>o<in>
<d6>	demonstrative agreement class 6	13	ku<locpf>la<d>w<d6>o<pos2>ma<n6>dwala<nr>
<d7>	demonstrative agreement class 7	29	a<z6>nga<p>le<d>s<d7>o<pos2>si<n7>khathi<nr>
<d9>	demonstrative agreement class 9	54	ba<z2>le<d9>
<d10>	demonstrative agreement class 10	17	ku<i15>yi<p>le<d>z<d10>o<pos2>
<d11>	demonstrative agreement class 11	8	sa<z7>lo<d>lu<d11>
<d14>	demonstrative agreement class 14	5	lo<d>b<d14>o<pos2>
<d15>	demonstrative agreement class 15	26	i<i9>lo<d>kh<d15>o<pos2>
<der>	derivational morpheme	70	a<iv>ba<n2>zal<vr>a<der>
<dim>	diminutive	106	a<iv>ma<n6>dod<nr>ana<dim>
<fem>	feminine	50	a<iv>ma<n6>doda<nr>kazi<fem>
<fut>	futurative	571	a<hort>si<s1p>zo<fut>satsh<vr>is<xc>w<xp>a<va>
<g1>	negative subject prefix class 1	60	a<neg>k<g1>azi<vr>
<g1p>	negative subject prefix first person plural	14	a<neg>s<g1p>azi<vr>ke<ke>
<g1s>	negative subject prefix first person singular	59	a<neg>ng<g1s>az<vr>anga<vpg>
<g2>	negative subject prefix class 2	27	a<neg>b<g2>azi<vr>
<g2p>	negative subject prefix second person plural	2	a<neg>ni<g2p>se<asp>na<p>yo<pr9>
<g2s>	negative subject prefix second person singular	23	a<neg>kw<g2s>azi<vr>
<g3>	negative subject prefix class 3	2	a<neg>wu<g3>sa<asp>lind<vr>i<vg>
<g4>	negative subject prefix class 4	6	a<neg>yi<g4>khulum<vr>i<vg>
<g5>	negative subject prefix class 5	7	a<neg>li<g5>bon<vr>i<vg>
<g6>	negative subject prefix class 6	9	a<neg>w<g6>azi<vr>
<g7>	negative subject prefix class 7	7	a<neg>s<g7>azi<vr>
<g9>	negative subject prefix class 9	26	a<neg>yi<g9>balulek<vr>ile<vpl>
<g10>	negative subject prefix class 10	9	a<neg>zi<g10>fan<vr>i<vg>
<g11>	negative subject prefix class 11	2	a<neg>lu<g11>sh<vr>i<vg>
<g15>	negative subject prefix class 15	56	a<neg>ku<g15>kheth<vr>i<vg>
<hort>	hortative	109	a<hort>k<s1>enz<vr>e<vs>
<hum>	human	83	a<iv>ba<n2>dlal<vr>i<hum>
<i1>	indicative subject prefix class 1	964	i<iv>zi<n10>nyawo<nr>w<i1>a<asp>fik<vr>a<va>
<i1p>	indicative subject prefix first person plural	246	a<r>s<i1p>a<asp>shiy<vr>a<va>
<i1s>	indicative subject prefix first person singular	496	a<r>ngi<i1s>land<vr>iw<xp>e<vps>
<i2>	indicative subject prefix class 2	497	a<neg>ba<i2>nga<asp>thand<vr>a<va>
<i2p>	indicative subject prefix second person plural	38	e<r>n<i2p>a<asp>ni<p2p>m<o1>biz<vr>a<va>
<i2s>	indicative subject prefix second person singular	520	o<r>w<i2s>a<asp>b<vr>e<vps>
<i3>	indicative subject prefix class 3	294	a<r>w<i3>a<asp>zang<vr>e<vps>
<i4>	indicative subject prefix class 4	255	e<r>y<i4>a<asp>hluk<vr>en<xr>e<vps>
<i5>	indicative subject prefix class 5	164	e<r>l<i5>a<asp>li<o5>ngen<vr>e<vps>
<i6>	indicative subject prefix class 6	369	a<i6>b<past>e<p6>m<o1>gwem<vr>a<va>
<i6_vr>	indicative subject prefix class 6 + verb root	2	az<i6_vr>iw<xp>a<va>
<i7>	indicative subject prefix class 7	154	a<r>s<i7>ethus<vr>e<vps>

<i9>	indicative subject prefix class 9	343	a<r>y<i9>a<asp>buzis<vr>an<xr>a<va>
<i10>	indicative subject prefix class 10	209	a<r>zi<i10>fis<vr>a<va>yo<rsf>
<i11>	indicative subject prefix class 11	49	e<p1>yo<fut>lu<i11>bek<vr>a<va>
<i14>	indicative subject prefix class 14	29	b<i14>a<asp>b<vr>a<va>
<i15>	indicative subject prefix class 15	563	a<r>ku<i15>khumbul<vr>e<vps>
<imp>	imperative	67	bamb<vr>a<imp>
<in>	inanimate	246	a<i6>n<p>e<iv>si<n7>sind<vr>o<in>
<int>	interrogative	80	a<i6>thi<vr>ni<int>
<intj>	interjection	24	awu<intj>
<iv>	initial vowel of the noun	1785	a<i6>n<p>e<iv>si<n7>sind<vr>o<in>
<iv_n1>	initial vowel + noun class 1	192	be<p2>n<p>o<iv_n1>si<n7>khuni<nr>
<iv_n2>	initial vowel + noun class 2	25	ba<i2>be<past>ng<p>o<iv_n2>bani<nr>
<iv_n3>	initial vowel + noun class 3	48	e<r>si<i7>n<p>o<iv_n3>bhanana<nr>
<iv_n5>	initial vowel + noun class 5	245	a<i6>y<p>i<iv_n5>shumi<nr>
<iv_n9>	initial vowel + noun class 9	40	e<p1>y<p>i<iv_n9>nyanga<nr>
<iv_n11>	initial vowel + noun class 11	96	e<p1>n<p>o<iv_n11>laka<nr>
<iv_n14>	initial vowel + noun class 14	7	n<p>o<iv_n14>bovu<nr>
<ke>	the suffix ke	9	a<neg>s<g1p>azi<vr>ke<ke>
<locpf>	locative prefix	540	a<i6>se<locpf>zi<n10>sefw<nr>eni<locsf>
<locsf>	locative suffix	247	a<i6>se<locpf>zi<n10>sefw<nr>eni<locsf>
<mr>	modal/auxiliary root	160	a<hort>wu<s2s>zange<mr>
<n1>	noun class 1	226	a<neg>ka<g1>se<asp>mu<n1>bi<ar>
<n2>	noun class 2	107	a<iv>ba<n2>dlal<vr>i<hum>
<n3>	noun class 3	228	a<iv>ma<n6>n<n9>hlak<nr>o<iv>mu<n3>zi<nr>
<n4>	noun class 4	85	a<i6>y<p>i<iv>mi<n4>ningi<ar>
<n5>	noun class 5	23	ba<i2>li<n5>shum<nr>
<n6>	noun class 6	297	a<i6>ye<p6>se<mr>ma<n6>ningi<ar>
<n7>	noun class 7	293	a<i6>n<p>e<iv>si<n7>sind<vr>o<in>
<n9>	noun class 9	445	a<i6>ye<p6>n<p>e<iv>n<n9>tukuthel<vr>o<in>
<n10>	noun class 10	324	a<i6>n<p>e<iv>zi<n10>mali<nr>
<n11>	noun class 11	26	a<neg>ngi<g1s>lu<n11>tho<nr>
<n14>	noun class 14	63	b<z14>o<iv>bu<n14>gazagaza<nr>
<n15>	noun class 15	551	a<hort>wu<s2s>ku<n15>zw<vr>e<vs>
<neg>	negative	689	a<i6>nga<neg>ka<asp>y<vr>a<va>
<nr>	noun root	2590	a<i6>n<p>e<iv>zi<n10>mali<nr>
<o1>	object prefix class 1	470	a<hort>ngi<s1s>m<o1>gijim<vr>el<xa>e<vs>
<o1p>	object prefix first person plural	59	a<i6>b<past>e<p6>si<o1p>bek<vr>e<vps>
<o1s>	object prefix first person singular	170	a<hort>ku<s15>ngi<o1s>ph<vr>e<vs>
<o2>	object prefix class 2	105	a<i6>ba<o2>bon<vr>a<va>
<o2p>	object prefix second person plural	10	ku<i15>ni<o2p>phath<vr>e<vps>
<o2s>	object prefix second person singular	94	a<i6>sa<asp>ku<o2s>lind<vr>ile<vpl>
<o3>	object prefix class 3	86	a<i6>wu<o3>nyakaz<vr>is<xc>a<va>
<o4>	object prefix class 4	156	a<i6>y<o4>enz<p1_vr>e<vps>
<o5>	object prefix class 5	74	a<i6>li<o5>bhek<vr>isis<xi>a<va>
<o6>	object prefix class 6	75	a<i6>wa<o6>thol<vr>e<vps>
<o7>	object prefix class 7	47	a<i6>si<o7>bek<vr>el<xa>a<va>
<o9>	object prefix class 9	236	a<hort>si<s1p>yi<o9>yek<vr>e<vs>
<o10>	object prefix class 10	201	a<hort>si<s1p>z<o10>ethwel<vr>e<vs>
<o11>	object prefix class 11	19	a<i6>lu<o11>pheth<vr>e<vps>
<o14>	object prefix class 14	15	a<i6>bu<o14>vez<vr>e<vps>
<o15>	object prefix class 15	226	a<i6>be<past>ku<o15>cel<vr>e<vps>
<opt>	optative	5	a<i6>bo<opt>buy<vr>a<va>
<p>	preposition	1405	a<i6>n<p>e<iv>si<n7>sind<vr>o<in>
<p1>	participial subject prefix class 1	517	e<p1>b<vr>e<vps>
<p1_vr>	participial subject prefix class 1 + verb root	130	a<i6>sebe<vr>sebenz<p1_vr>a<va>
<p1p>	participial subject prefix first person plural	89	be<past>si<p1p>m<o1>lind<vr>el<xa>e<vps>
<p1s>	participial subject prefix first person singular	227	be<past>ng<p1s>azi<vr>
<p2>	participial subject prefix class 2	178	a<r>b<i2>a<asp>be<p2>jayiv<vr>a<va>
<p2p>	participial subject prefix second person plural	18	e<r>n<i2p>a<asp>ni<p2p>m<o1>biz<vr>a<va>
<p2s>	participial subject prefix second person singular	194	o<r>s<asp>u<p2s>bhal<vr>iw<xp>e<vps>
<p3>	participial subject prefix class 3	95	se<asp>wu<p3>qal<vr>e<vps>
<p4>	participial subject prefix class 4	83	i<p4>b<vr>e<vps>
<p5>	participial subject prefix class 5	91	be<past>li<p5>kade<mr>
<p6>	participial subject prefix class 6	170	a<i6>b<past>e<p6>m<o1>gwem<vr>a<va>
<p7>	participial subject prefix class 7	48	be<past>se<asp>s<p7>enz<vr>ek<xn>a<va>
<p9>	participial subject prefix class 9	142	be<past>yi<p9>khulum<vr>a<va>
<p10>	participial subject prefix class 10	113	a<r>zi<p10>fis<vr>a<va>yo<rsf>
<p11>	participial subject prefix class 11	21	be<past>lu<p11>qond<vr>e<vps>ni<int>
<p14>	participial subject prefix class 14	9	b<i14>a<asp>bu<p14>kho<adv>na<st>

<p15>	participial subject prefix class 15	307	be<past>ku<p15>lo<pr5>
<past>	past tense morpheme	169	a<i6>b<past>e<p6>m<o1>gwem<vr>a<va>
<pl>	(imperative) plural	10	dl<vr>a<imp>nini<pl>
<pos2>	position 2	78	a<z6>nga<p>le<d>s<d7>o<pos2>si<n7>khathi<nr>
<pos3>	position 3	22	b<z14>a<z6>ku<locpf>lo<d1>wa<st>ya<pos3>
<pot>	potential	14	a<s6>nga<pot>vum<vr>a<va>
<pr1>	pronoun class 1	39	a<neg>ka<g1>na<p>ye<pr1>
<pr1p>	pronoun first person plural	34	a<iv>ba<n2>fo<nr>w<z1>ethu<pr1p>
<pr1s>	pronoun first person singular	32	a<z6>mi<pr1s>
<pr2>	pronoun class 2	34	a<i6>ye<p6>ku<locpf>bo<pr2>
<pr2p>	pronoun second person plural	7	ki<locpf>ni<pr2p>
<pr2s>	pronoun second person singular	26	a<neg>ku<g15>kwa<p>kho<pr2s>
<pr3>	pronoun class 3	16	a<i6>ye<p6>se<mr>na<p>wo<pr3>
<pr4>	pronoun class 4	4	la<z5>yo<pr4>
<pr5>	pronoun class 5	23	a<i6>ye<p6>yi<p>lo<pr5>
<pr6>	pronoun class 6	6	a<i6>wo<pr6>dwa<qr>
<pr7>	pronoun class 7	10	i<iv_n5>wa<z1>so<pr7>
<pr9>	pronoun class 9	38	a<i6>ye<p6>ku<locpf>yo<pr9>
<pr10>	pronoun class 10	18	a<neg>ku<g15>zo<pr10>na<st>
<pr11>	pronoun class 11	2	lo<pr11>na<st>
<pr14>	pronoun class 14	7	bo<pr14>na<st>
<pr15>	pronoun class 15	13	a<neg>ku<g15>yi<p>kho<pr15>
<pres>	presentative	16	na<p>se<locpf>baba<nr>nan<pres>g<prs1>o<pos2>
<prs1>	presentative agreement class 1	4	nan<pres>gu<prs1>ya<pos3>
<prs2>	presentative agreement class 2	1	nam<pres>pa<prs2>
<prs5>	presentative agreement class 5	2	nan<pres>t<prs5>o<pos2>
<prs6>	presentative agreement class 6	1	nan<pres>ka<prs6>
<prs7>	presentative agreement class 7	1	na<pres>si<prs7>
<prs9>	presentative agreement class 0	1	nan<pres>si<prs9>
<prs10>	presentative agreement class 10	3	na<pres>z<prs10>o<pos2>
<prs15>	presentative agreement class 15	3	na<pres>kh<prs15>o<pos2>
<qr>	quantifier root	31	a<i6>wo<pr6>dwa<qr>
<r>	relative morpheme	1205	a<r>ma<n6>hlanu<ar>
<red>	reduplication	31	a<iv>ma<n6>ngele<red>ngele<nr>
<refl>	reflexive object prefix	78	a<hort>si<s1p>zi<refl>yek<vr>e<vs>
<rsf>	relative suffix	146	a<r>b<i2>a<asp>ku<o15>fis<vr>a<va>yo<rsf>
<s1>	subjunctive subject prefix class 1	167	a<hort>k<s1>enz<vr>e<vs>
<s1p>	subjunctive subject prefix first person plural	55	a<hort>si<s1p>bong<vr>e<vs>
<s1s>	subjunctive subject prefix first person singular	80	a<hort>ngi<s1s>bang<vr>e<vs>
<s2>	subjunctive subject prefix class 2	86	a<r>ba<s2>fund<vr>e<vs>
<s2p>	subjunctive subject prefix second person plural	14	ni<s2p>b<vr>e<vs>
<s2s>	subjunctive subject prefix second person singular	80	a<hort>wu<s2s>bang<vr>e<vs>
<s3>	subjunctive subject prefix class 3	25	a<hort>wu<s3>gcob<vr>e<vs>
<s4>	subjunctive subject prefix class 4	19	i<s4>b<vr>e<vs>
<s5>	subjunctive subject prefix class 5	15	a<hort>li<s5>phakam<vr>is<xc>e<vs>
<s6>	subjunctive subject prefix class 6	96	a<hort>wa<s6>sal<vr>e<vs>
<s6_vr>	subjunctive subject prefix class 6 + verb root	1	and<s6_vr>is<xc>e<vs>
<s7>	subjunctive subject prefix class 7	23	a<hort>si<s7>hamb<vr>e<vs>
<s9>	subjunctive subject prefix class 9	34	a<hort>yi<s9>bhek<vr>e<vs>
<s10>	subjunctive subject prefix class 10	24	a<hort>zi<s10>bulal<vr>e<vs>
<s11>	subjunctive subject prefix class 11	2	lu<s11>b<vr>e<vs>
<s14>	subjunctive subject prefix class 14	2	bu<s14>thul<vr>e<vs>
<s15>	subjunctive subject prefix class 15	60	a<hort>ku<s15>b<vr>e<vs>
<st>	stabilizer	114	a<neg>ka<g1>bo<pr2>na<st>
<va>	verb suffix	4768	a<hort>k<s1>ezw<vr>a<va>
<vg>	negative verb suffix	317	a<i6>ye<p6>nga<neg>sa<asp>phum<vr>i<vg>
<voc>	vocative prefix	1	we<voc>ba<n2>ntu<nr>
<vpg>	negative perfect verb suffix	42	a<i6>ye<p6>nga<neg>khish<vr>w<xp>anga<vpg>
<vpl>	perfect verb suffix long form	436	a<i6>hlakaniph<vr>ile<vpl>
<vps>	perfect verb suffix short form	1763	a<i6>b<past>e<p6>si<o1p>bek<vr>e<vps>
<vr>	verb root	8544	a<hort>k<s1>enz<vr>e<vs>
<vs>	subjunctive verb suffix	642	a<hort>k<s1>enz<vr>e<vs>
<w>	word	1	kw<i15>a<asp>shay<vr>a<va>n<n9>sumo<w>ni<int>
<xa>	applied extension	626	a<hort>kw<s15>enz<vr>el<xa>e<vs>
<xc>	causative extension	519	a<hort>li<s5>phakam<vr>is<xc>e<vs>
<xi>	intensive extension	19	a<i6>cabang<vr>isis<xi>e<vps>
<xn>	neuter extension	345	a<i6>bon<vr>akal<xn>e<vps>
<xp>	passive extension	766	a<hort>si<s1p>zo<fut>satsh<vr>is<xc>w<xp>a<va>
<xr>	reciprocal extension	163	a<i6>nga<asp>m<vr>el<xa>an<xr>a<va>

<z>	possessive morpheme	2	m<n1>ngan<nr>a<z>mi<pr1s>
<z1>	possessive agreement class 1	118	a<iv>ba<n2>fo<nr>w<z1>ethu<pr1p>
<z2>	possessive agreement class 2	46	a<iv>ba<n2>nta<nr>b<z2>a<iv>ba<n2>ntu<nr>
<z3>	possessive agreement class 3	42	ka<z3>
<z4>	possessive agreement class 4	61	ka<z4>
<z5>	possessive agreement class 5	112	e<r>la<z5>kwa<p>mbabo<nr>
<z6>	possessive agreement class 6	27	a<r>w<z6>e<iv_n5>bhubesi<nr>
<z6_iv>	possessive agreement class 6 + initial vowel	19	a<z6_iv>ma<n6>futha<nr>
<z6_iv_n1>	possessive agreement class 6 + initial vowel + noun class 1	1	o<z6_iv_n1>n<n9>kos<nr>az<fem>ana<dim>
<z6_iv_n3>	possessive agreement class 6 + initial vowel + noun class 3	1	o<z6_iv_n3>gwayi<nr>
<z7>	possessive agreement class 7	68	i<i9>nga<p>kw<z15>e<r>sika<z7>baba<nr>
<z9>	possessive agreement class 9	104	e<r>y<z9>e<iv>zi<n10>tha<nr>
<z10>	possessive agreement class 10	67	e<r>z<z10>e<iv>mi<n4>sebenz<vr>i<der>
<z11>	possessive agreement class 11	15	luka<z11>ma<der>m<n1>ncube<nr>
<z14>	possessive agreement class 14	16	b<z14>a<z6>ku<locpf>lo<d1>wa<st>ya<pos3>
<z15>	possessive agreement class 15	160	ba<i2>se<locpf>mva<adv>kwa<z15>mi<pr1s>

2 Noun classes in Zulu

There are twelve noun classes in Zulu. These classes are numbered 1–7, 9, 10, 11, 14, 15. Typically, the classes are identified by distinctive noun prefixes (the second prefix in the following words). However, some classes have members which lack the noun prefix; in the case of nouns of classes 5 and 11, the prefix is lacking completely. Class 2 has an alternate form where the initial vowel and the noun prefix have become fused. A lacking prefix is indicated by $\emptyset$ in the following examples below.

noun class	**example**	**noun class**	**example**
1	u-mu-ntu *'person'*	2	a-ba-ntu *'people'*
	u-$\emptyset$-baba *'father*		o-baba *'fathers'*
3	u-mu-zi *'village'*	4	i-mi-zi *'villages'*
	u-$\emptyset$-nogwaja *'hare'*	2	o-nogwaja *'hares'*
5	i-$\emptyset$-gama *'name'*	6	a-ma-gama *'names'*
7	i-si-tsha *'dish'*	10	i-zi-tsha *'dishes'*
9	i-m-pala *'impala'*	10	i-zim-pala *'impalas'*
	i-$\emptyset$-khwaya *'choir'*	6	a-ma-khwaya *'choirs'*
11	u-$\emptyset$-phondo *'horn'*	10	i-zim-pondo *'horns'*
14	u-bu-hle *'beauty'*		
	u-$\emptyset$-tshani *'grass'*		
15	u-ku-dla *'food'*		

The numbering system was devised by [1], and reflects the historical affinities between Zulu and other Bantu languages: Zulu lacks classes 8, 12 and 13, which are found in other Bantu languages. In the labels used on the database, morphemes that command or show agreement have been labeled as *<xn>*, where *x* is a letter or sequence of letters, and *n* is a number: thus the morpheme *m-* in *mfundi* is labeled *<n1>*, as it marks the noun as belonging to noun class *1*. The morpheme *si-* in *engisifundisile* is marked *<o7>*, as it shows object agreement with a noun of class *7*.

3 Part-of-speech tagging based on the morphological structure of a word

In this section we list a set of 34 rules which were provided by a linguistic expert and used to assign the part-of-speech (POS) tag to a word whose morphological structure is known. Zulu words can often be POS-tagged by their first morpheme alone. Sometimes, however, the correct tagging can only be found from a combination of the first morpheme (*morpheme1*), and one other morpheme, which may be referred to as the identifier (abbreviated *J*). There may be several *J* morphemes in a word, but it is always the leftmost one which identifies the word (*jMorpheme1*). The *J* morpheme may be separated from the first morpheme by several other morphemes. *J* morphemes are [<ar>, <adv>, <advpf>, <d>, <dX>, <fut>, <imp>, <locpf>, <mr>, <nr>, <nX>,

<opt>, <p>, <prX>, <qr>, <r>, <vr>]. Furthermore, we introduce the variable X which is a place holder for the following values: [1 - 15, 1s, 2s, 1p, 2p].

Remark: Elements in a list correspond to a disjunction (logic OR). For instance a rule

if morpheme1 = <a> and jMorpheme1 = [<b>, <c>] then label *x*, means

if morpheme1 is <a> and jMorpheme1 is <b> or <c> then the label is *x*.

Rules are in no particular order, however, the identification by first morpheme and first J morpheme has to be performed before the identification by first morpheme only.

3.1 Identification by first morpheme and first J morpheme

1. *if* morpheme1 = <red> and jMorpheme1 = <adv>

   → *adv (adverb)*

2. *else if* morpheme1 = <asp> and jMorpheme1 = [<adv>, <advpf>, <ar>, <locpf>, <nr>, <p>, <prX>, <r>, <pX>, <nX>]

   → *cop (copulative)*

3. *else if* morpheme1 = <iX> and jMorpheme1 = [<adv>, <advpf>, <ar>, <dX>, <in>, <locpf>, <nr>, <p>, <prX>, <r>, <nX>, <d>, <dX>]

   → *cop (copulative)*

4. *else if* morpheme1 = <neg> and jMorpheme1 = [<adv>, <advpf>, <ar>, <locpf>, <nr>, <p>, <prX>, <r>, <nX>]

   → *cop (copulative)*

5. *else if* morpheme1 = <past> and jMorpheme1 = [<adv>, <advpf>, <locpf>, <nr>, <p>, <prX>, <r>]

   → *cop (copulative)*

6. *else if* morpheme1 = <pX> and jMorpheme1 = [<adv>, <advpf>, <ar>, <locpf>, <nr>, <p>, <prX>, <r>, <nX>, <d>, <dX>]

   → *cop (copulative)*

7. *else if* morpheme1 = <st> and jMorpheme1 = [<ar>, <nX>]

   → *cop (copulative)*

8. *else if* morpheme1 = [<asp>, <iX>, <neg>, <past>, <pX>] and jMorpheme1 = <mr>

   → *m (modal)*

9. *else if* morpheme1 = <d> and jMorpheme1 = [<nX>, <nr>]

   → *n (noun)*

10. *else if* morpheme1 = <dX> and jMorpheme1 = [<nX>, <nr>]

    → *n (noun)*

11. *else if* morpheme1 = <vr> and jMorpheme1 = <in>

    → *n (noun)*

12. *else if* morpheme1 = <red> and jMorpheme1 = <nr>

→ *n (noun)*

13. *else if* morpheme1 = <prX> and jMorpheme1 = <st>

    → *pron (pronoun)*

14. *else if* morpheme1 = <prX> and jMorpheme1 = [<qr>, <nX>]

    → *q (quantifier)*

15. *else if* morpheme1 = [<asp>, <iX>, <neg>, <past>, <pX>] and jMorpheme1 = [<vr>, <fut>, <opt>]

    → *v (verb)*

16. *else if* morpheme1 = [<oX>, <red>, <refl>, <st>, <vr>] and jMorpheme1 = <imp>

    → *v (verb)*

17. *else if* morpheme1 = <vr> and jMorpheme1 = <pl>

    → *v (verb)*

18. *else if* morpheme1 <oX> and jMorpheme1 = <vr>

    → *v (verb)*

19. *else if* morpheme1 = [<red>, <refl>, <st>] and jMorpheme1 = <vr>

    → *v (verb)*

3.2 Identification by first morpheme only

20. *else if* morpheme1 = <ar> → *a (adjective)*
21. *else if* morpheme1 = [<adv>, <advpf>] → *adv (adverb)*
22. *else if* morpheme1 = <cj> → *conj (conjunction)*
23. *else if* morpheme1 = [<d>, <dX>] → *dem (demonstrative)*
24. *else if* morpheme1 = <intj> → *intj (interjection)*
25. *else if* morpheme1 = <locpf> → *loc (locative)*
26. *else if* morpheme1 = [<iv>, <iv_nX>, <nX>, <nr>, <der>, <voc>] → *n (noun)*
27. *else if* morpheme1 = [<zX>, <zX_iv>] → *pos (possessive)*
28. *else if* morpheme1 = <p> → *p (prepositional)*
29. *else if* morpheme1 = <pres> → *pres (presentative)*
30. *else if* morpheme1 = <prX> → *pron (pronoun)*
31. *else if* morpheme1 = <r> → *rel (relative)*
32. *else if* morpheme1 = [<hort>, <iX_vr>, <pX_vr>, <sX>] → *v (verb)*
33. *else if* only label = <w> → *w (word)*
34. *else* → *unknown*

These rules have been implemented into a prototype of a POS-tagger available from `http://www.cs.bris.ac.uk/Research/MachineLearning/Morphology/Resources/`.

References

[1] Carl Meinhof. *Grundzüge einer vergleichenden Grammatik der Bantusprachen*. Reimer, Berlin., 1906.

[2] Sebastian Spiegler, Andrew van der Spuy, and Peter A. Flach. Ukwabelana - an open-source morphological Zulu corpus. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, 2010.